

The BBN RT03 BN English System

Long Nguyen, Nicolae Duta, John Makhoul,
Spyros Matsoukas, Richard Schwartz,
Bing Xiang, Dongxin Xu

RT-03 Workshop

Boston, MA, May 19-21, 2003

Overview

- Recognition strategy
- Development test set
- Improvements
- Evaluation results

Recognition Strategy



- Decoding steps
 - Fast-match using PTM to save word-graphs
 - Viterbi search on word-graphs using SCTM to output N-bests
 - N-bests rescoring using cross-word SCTM and 4-grams
- Decode three times with adaptation in between

3

BBN TECHNOLOGIES
A Verizon Company

Development Test Set Creation



- BBN+LIMSI Team initiated the effort; CU and SRI joined
- Six episodes from all sources in the TDT4 English corpus
 - Broadcast in the second half of Jan '01
 - First 30 minutes from each episode (~3 hours)

	% WER
0. Baseline (RT-02 Mothballed system)	16.3

4

BBN TECHNOLOGIES
A Verizon Company

Audio Automatic Segmentation



- Segmentation procedure
 - Separate stream into wide- and narrow-band chunks
 - Detect gender changes and pauses in each band
 - Detect speaker changes within a band+gender turn
 - Cluster speaker turns (new)
 - Chop into short segments (~ 4 seconds on average)

	% WER
0. Baseline (RT-02 Mothballed system)	16.3
1. Baseline + New auto segmentation	16.1

5

BBN TECHNOLOGIES
A Verizon Company

Importing CTS Technologies



- Imported relevant technologies developed for CTS in the last two years
 - Modern ML training (fuzzy labels, PLP, HLDA, CMLLR-SAT, GI, and no band-specific modeling)
 - HLDA-SAT (with speaker turn clustering)
 - MMI training

	% WER
1. Baseline + New auto segmentation	16.1
2. Modern ML training	15.2
3. + HLDA-SAT	14.6
4. + MMI training	13.7

6

BBN TECHNOLOGIES
A Verizon Company

4-Gram Language Model



- 60k-word lexicon (0.68% OOV rate)
- Used all H4 LM training data (1046M words)
 - Pool with weights and count (keeping all)
 - 250M 3-grams, 82% hit rate, perplexity = 178
 - 580M 4-grams, 58% hit rate, perplexity = 161

	%WER
4. + MMI training	13.7
5. + 4-gram rescoreing	12.9

Use of TDT4 Data



- Lightly-supervised decoding to extract data
 - Add TDT4 closed-captions into LMs
 - Decode and extract 73 hours of ‘utterances’ with zero WER
- Pooled (140h H4 + 73h TDT4) and retrained

	%WER
5. + 4-gram rescoreing	12.9
6. + TDT4 acoustic training data	11.9

The 10xRT System



- Used Fast Gaussian Computation (FGC)
- Used narrow pruning beam-widths (both in decoding and adaptation)
- Sped-up system ran at 5.5xRT on an Intel Xeon 2.8GHz machine

	%WER
6. + TDT4 acoustic training data	11.9
7. + Speedup options to run < 10xRT	12.2

Updated Language Models



- 62k-word lexicon (0.35% OOV rate)
- Used more language training data (+1G words)
 - GigaWord (upto Jan 15, 2001)
 - TDT2 and TDT4 (upto Jan 15, 2001)
 - CNN transcripts (Jan 01, 2000 – Jan 15, 2001)
- 710M 4-grams: 61% hit rate, perplexity = 139

	%WER
7. + Speedup options to run < 10xRT	12.2
8. + Updated Lexicon and LMs	11.8

One More Pass of Adaptation



- System ran too fast (16h:23m or 5.5xRT)
 - Segmentation: 0h:18m
 - Un-adapted decoding: 5h:58m
 - Adapted decoding: adapt: 4h:32m + decode: 5h:35m
- Added one more pass of adaptation
 - Adapt: 0h:30m + Decode: 5h:35m

	% WER
8. + Updated Lexicon and LMs	11.8
9. + One more pass of adaptation	11.6

11

BBN TECHNOLOGIES
A Verizon Company

Roadmap of Improvements



	% WER
0. Baseline (RT-02 mothballed system)	16.3
1. Baseline + New auto segmentation	16.1
2. Modern ML training	15.2
3. + HLDA-SAT	14.6
4. + MMI training	13.7
5. + 4-gram rescoring	12.9
6. + TDT4 acoustic training data	11.9
7. + Speedup options to run < 10xRT	12.2
8. + Updated lexicon and LMs	11.8
9. + One more pass of adaptation	11.6

12

BBN TECHNOLOGIES
A Verizon Company

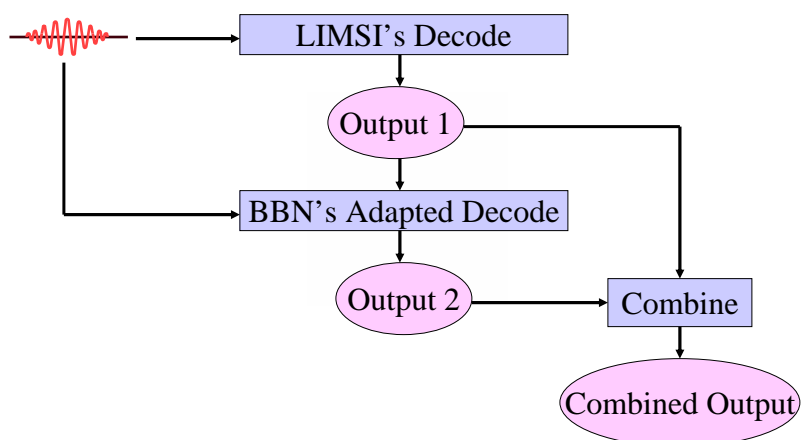
System Combination



- Combined final results produced by BBN's and LIMSI's systems

	% WER	xRT
BBN	11.6	7.5
LIMSI	11.8	9.8
BBN+LIMSI Combination	10.3	17.3

Post-Eval Combination



Benchmark Results



	Dev03	Eval03	Prog03
RT-02	16.3	?	18.0
RT-03	11.6	11.2	13.8
Relative WER Reduction	28.2%	?	23.3%

Summary



- Achieved substantial WER reduction for both the Development and the Progress test sets
- Tech transfer from CTS to BN worked well
- Took advantage of large amounts of training data (both audio and text)